INTELLIGENT CLASSIFICATION OF ELECTROLARYNGOGRAPH SIGNALS

RT Ritchings[1], M.McGillion[1] CJ Moore[2]
[1]Computer Science, School of Sciences, University of Salford, UK
[2]North Western Medical Physics, Christie Hospital NHS Trust, Manchester, UK

***Abstract*-This paper describes a prototype system for the intelligent classification of electrolaryngograph (EGG) signals in order to provide an objective assessment of voice quality in patients at different stages of recovery after treatment for larynx cancer. The system extracts salient short-term and long-term time-domain and frequency-domain parameters from EGG signals taken from male patients steadily phonating the vowel /i/. The quality of these voices was also independently assessed by a Speech and Language Therapist (SALT) according to their 7-point ranking of subjective voice quality. These data were used to train and test a Multi-layer Perceptron (MLP) neural network to classify EGG signals in terms of voice quality. Several MLP configurations were investigated using various combinations of these signal parameters, and the best results were obtained using a combination of short-term and long-term parameters, for which an accuracy of 92% was achieved. It is envisaged that this system could be used as a valuable aid to the SALT during clinical evaluation of voice quality.**

***Keywords* – Intelligent Systems, MLP neural networks, electrolaryngograph signals, voice quality, larynx cancer**

I. INTRODUCTION

Electrolarygography (EGG) measures the impedance of the electrical signals transmitted through the living tissue surrounding the larynx in the neck [1]. Earlier work has shown that there are structural differences in the EGG signals of stationary vowels for normal and pathological voices (Fig.1), and that parameters derived from the EGG signals can be used to train a Multi-layer Perceptron (MLP) neural network to classify the signals as normal or pathological with an accuracy of 80% [2].

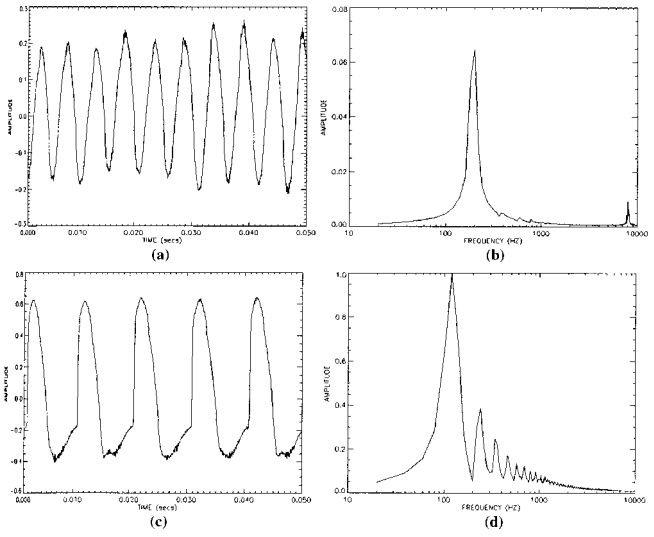

Fig. 1. Electrolaryngogaph signals and frequency spectrum for males phonating /i/: pathological voice a) and b), normal voice c) and d)

Whilst this system provided good classification between normal and abnormal voice quality, the feature set was limited to sub-optimal classification results as it is well known that some pathologies are measured more easily using long-term (>50ms) parameters [3]. This paper describes the refinement of the MLP approach to objective voice quality assessment, by introducing long-term features to the classification system.

In addition, the system has been extended to provide a graded classification of pathological voices in line-with the 7-point ranking scheme used by Speech and Language Therapists (SALT) to assess voice quality. At present, SALTs endeavor to rehabilitate a patient's voice back to normality, or as near normal as possible, quickly following treatment. Their ranking scheme (0=least abnormal, 6=most abnormal) is based on a variety of sound parameters, some of which are well defined, such as shimmer and jitter, while others, such as whisper and creak are descriptive or have tenuous physical correlates, As a result, the assessment is largely subjective and depends upon the experience of the SALT. The intelligent voice quality system described here aims to improve this situation by providing accurate, reproducible, graded measures of a patient's voice quality to help the SALT plan the patient's rehabilitation more accurately.

II. DATA CAPTURE

The data used to develop the system was captured off-line under clinical conditions at the Christie and Withington Hospitals in Manchester, using an Electrolaryngograph PCLX system. This system is used to capture the electrolaryngograph signals using pads placed either side of the neck. Acoustic signals were also recorded using a microphone, but have not been used at this stage. Both the EGG and acoustic data channels were captured synchronously at 20kHz with 16-bit Analog-to-Digital converters for up to 3 seconds while the subject phonated the vowel /i/ as steadily as possible.

Although speech data was recorded for both male and female patients, the largest pathological group was male, so it is these speech signals that were used in the study, For each patient the SALT made a subjective voice quality assessment using a 7-point ranking.

Report Documentation Page

Report Date	Report Type	Dates Covered (from... to)
25OCT2001	N/A	-

Title and Subtitle Intelligent Classification of Electrolaryngograph Signals	**Contract Number**
	Grant Number
	Program Element Number
Author(s)	**Project Number**
	Task Number
	Work Unit Number
Performing Organization Name(s) and Address(es) Computer Science, School of Sciences, University of Salford, UK	**Performing Organization Report Number**
Sponsoring/Monitoring Agency Name(s) and Address(es) US Army Research Development & Standardization Group (UK) PSC 803 Box 15 FPO AE 09499-1500	**Sponsor/Monitor's Acronym(s)**
	Sponsor/Monitor's Report Number(s)

Distribution/Availability Statement
Approved for public release, distribution unlimited

Supplementary Notes
Papers from the 23rd Annual International conference of the IEEE Engineering in Medicine and Biology Society, October 25-28, 2001, held in Istanbul, Turkey. See also ADM001351 for entire conference on cd-rom., The original document contains color images.

Abstract

Subject Terms

Report Classification unclassified	**Classification of this page** unclassified
Classification of Abstract unclassified	**Limitation of Abstract** UU

Number of Pages
4

III. Data processing

A voicing analysis was performed upon each 3-second EGG signal to determine if the subject had voiced during phonation. Voicing occurs when the vocal folds are vibrating, and as a result, the signal contains a fundamental frequency, f_0. If the signal was considered to be voiced, it was initially processed to extract the long-term features, and then the short-term features for classification of voice quality. The long-term features are mean fundamental frequency, Mf_0, standard deviation of f_0, SDf_0, and percentage of the signal that is voiced, V+, while the short-term features include parameters related to the structure of the first few harmonics, and the glottal noise.

The voicing test involved taking 50msec frames from the signals and applying Cepstral analysis techniques [4], to identify the voiced frames. Each frame was then pre-emphasised by forward differencing to suppress the effects of drifting signal amplitude, and its autocovariance multiplied by a Tukey-Hanning window, prior to transformation to the frequency domain using the Fast Fourier Transform. An estimate of f_0 for each frame, deduced during the voicing analysis, was used to derive the FHN normalised spectral representation [5]. This process removed any inter-patient variability in f_0 and its harmonics allowing a more effective modelling of the spectral envelope among groups of patients. Once the FHN spectrum had been determined, Gaussians were fitted to the data around f_0 and its first few harmonics. Each Gaussian, G_h, (h=0 up to typically 8) was parameterised as:

$$G_h = (\text{position}_h, \text{width}_h \text{ and amplitude}_h)$$

An observation was made that the mixture of Gaussians gave a better 'fit' to the FHN spectrum for the less abnormal patients, and so a parameter related to goodness of fit, called the Harmonic Linearity Measure (HLM), was calculated for each frame. Finally, as Glottal noise is considered to be an important measure of voice quality, a parameter, FHNNE, based on the Normalised Noise Energy (NNE) [6], but derived from the FHN spectrum, was calculated for the data.

The parameters extracted from the EGG signals and used for the MLP classification tests comprised of 3 long-term parameters (Mf_0, SDf_0, V+) and 17 short-term parameters (G_1, G_2, G_3, G_4, G_5, HLM, FHNNE). Full details of the data processing and extraction of these parameters can be found in McGillion [7].

IV. Data classification

A total number of 77 pathological EGG signals were available for training and testing data. For each of the 7 classes, 450 patterns were used for training/validation and 200 for testing. Unfortunately, as a result of the relatively small dataset, there were different numbers of patients in each class. As it is desirable to have equal numbers in each class to train an MLP adequately, additional frames were taken from some patients and a small percentage of the data was artificially generated by adding normally distributed noise to the short-term features of the existing patterns within each class.

A two-layer 7-output MLP was trained using the back-propagation training algorithm, softmax activation function, and cross-entropy error function. The advantage of using the softmax activation function, was that the output across all seven classes sums to 1.0 and can therefore be interpreted as a probability of membership of each of the seven classes, assuming equal prior probabilities. A further constraint placed upon the MLP is that for any single class to be declared the 'winner' the output for that class must be greater than 50% (0.5). MLP structures with different numbers of hidden units and subsets of the 21 input parameters were investigated in order to determine the combination that provided the minimum classification error.

V. Results and discussion

An overview of the best results obtained from the different combinations of input parameters and hidden units shown in Table 1.

The *best input set* is the MLP structure whose input parameters provided the best classification results, regardless of weight initialisation and the number of hidden units. The *best individual structure* is the MLP that provided the best classification results taking into account the number of hidden units in the MLP, but disregarding the weight hidden

TABLE 1
TEST RESULTS FOR THE 7-CLASS MLP

	Accuracy (%)	SD	Sensitivity (%) / Specificity (%)							Structure	Inputs
			Class 0	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6		
Best Individual MLP	92.00	6.42	98.5	96.0	94.5	80.5	86.5	91.5	96.5	20-25-7	G_1,G_2,G_3,G_4,G_5, FHNNE, HLM, Mf_0, SDf_0,V+
			0.12	0.87	0.75	4.12	3.0	2.12	0.75		
Best Individual Structure	90.30	1.65	98.1	94.6	93.1	86.9	83.6	81.7	94.1	20-40-7	G_1,G_2,G_3,G_4,G_5, FHNNE, Mf_0, SDf_0,V+
			0.47	1.3	1.42	3.1	3.82	4.17	1.37		
Best Input Set	87.24	3.47	98.0	93.5	92.2	81.1	77.3	74.8	93.5	21-[15,25,40]-7	G_1,G_2,G_3,G_4,G_5, f_0, FHNNE,HLM, Mf_0, SDf_0,V+
			0.45	1.49	1.52	4.02	4.67	5.28	1.25		

hidden units in the MLP, but disregarding the weight initialisations. The *best individual MLP* is that which takes into account both the number of hidden units and the weight initialisation.

The *best individual MLP* can be different from the *best individual structure* and the *best input set* if the variance in the classification ability of a given MLP is large (due to the MLP might produce a very high accuracy, the average

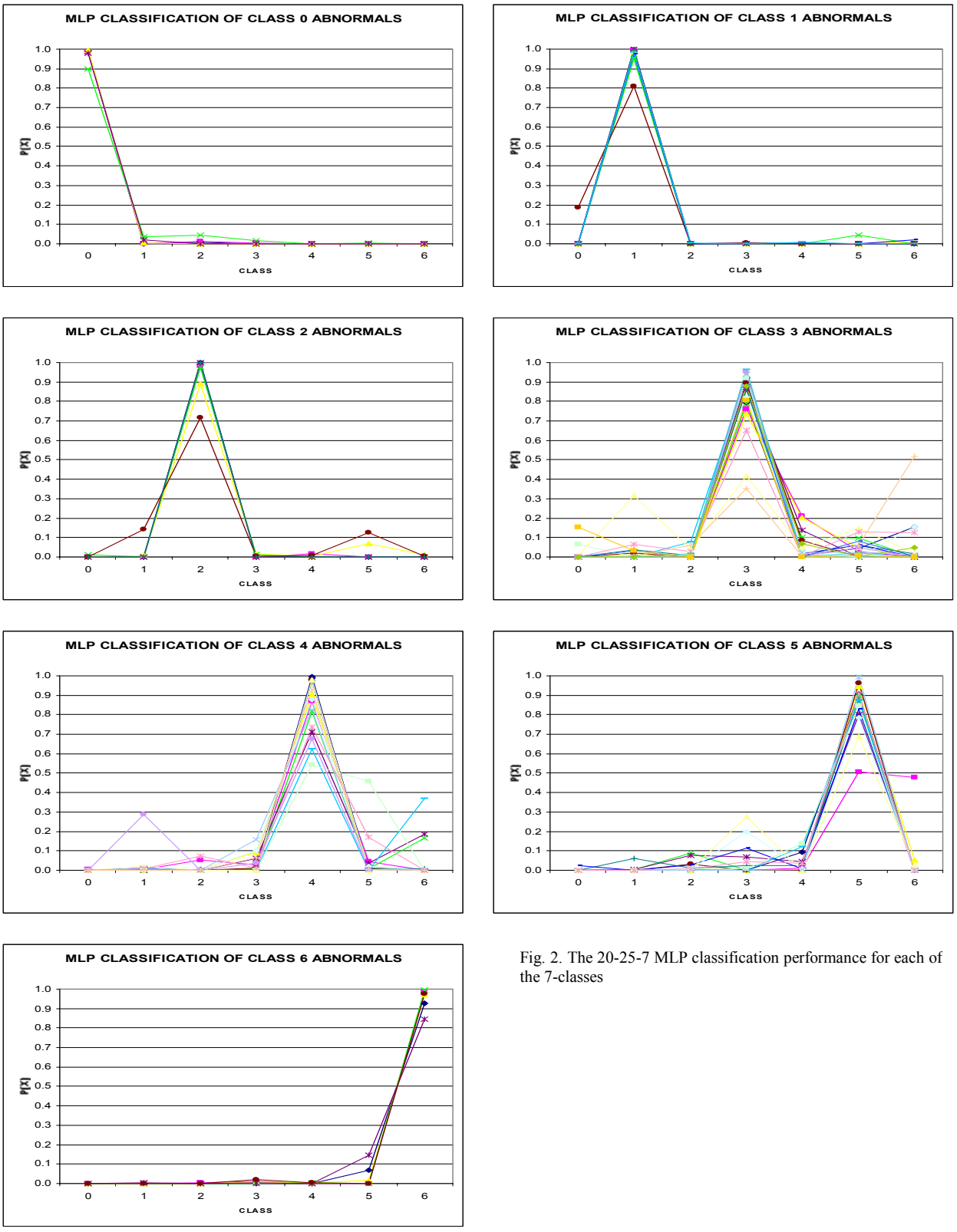

Fig. 2. The 20-25-7 MLP classification performance for each of the 7-classes

performance can be very poor. Therefore, it is important to consider all three results when determining the best performing input set and MLP structure.

The best overall MLP structure was a 20-25-7, using the parameters [G_1, G_2, G_3, G_4, G_5, FHNNE, HLM, Mf_0, SDf_0, V+], and the results indicate that this MLP was able to distinguish between the seven abnormal groups with an accuracy of up to 92%.

Figure 2 shows the output of the MLP for each pathological class. The output of the MLP is an estimate of the posterior probability of membership for each class. It should be noticed that there are only two cases where the output falls below the majority threshold 0.5, and only one misclassification (for class 3). Perhaps unsurprisingly, the classes at the two extremes of the scale, 0 and 6, provide the best classification results. In all cases, classes 3, 4, and 5 are the most difficult to discriminate between.

All the short-term features were found to contribute to the classification. The classification accuracy increased from 26.5% with [G_1] alone to 67.7% with [G_1, G_2, G_3, G_4, G_5]. Adding the other short-term features [FHNNE] and [HLM] increased the discrimination ability of the MLP to 72.07% and 68.64% respectively. Similarly, the long-term features, were also found to be very important to the discrimination between the classes. These parameters [MF0], [SDF0], [V+] alone were able to distinguish between the classes with accuracies of 37.57%, 23.07%, and 26.78% respectively. However, as can be seen from the results, it is the combination of the short-term and long-term features that provide the most accurate classifications of the abnormal signals.

VI. Conclusion

The results from this work suggest that a intelligent voice quality assessment system incorporating an MLP neural network can be trained to provide objective classifications of voice quality in line-with the 7-point ranking scheme used by the SALT.

However, it should noted that the MLP has been trained on the assessments of one SALT, which could lead to subjectively biased results. The collection of patient speech data, including voice quality rankings from several SALTs in the region is now taking place, and will hopefully provide a larger, and less biased dataset for training the system.

At the same time, work is taking place to identify and evaluate other parameters that can be derived from the speech data, in particular the acoustic data which has been largely ignored in this study so far, in order to further improve the accuracy and reproducibility of the system.

Acknowledgment

The support of this work by the EPSRC award GR/L51546 is greatly appreciated.

References

[1]Fourcin AJ, Abberton E, Miller D, Howell D (1996) *Laryngograph: Speech pattern element tools for therapy, training and assessment*. European Journal of Disorders of Communication 30(2)101-115

[2]Ritchings RT, McGillion M, Conroy G, Moore C (1999) *Objective assessment of pathological voice quality*. Proc IEEE SMC99, 2:340-345. ISBN: 0-7803-5683-7

[3]Baken RJ (1992) *Electroglottography*. Journal of Voice 6(2):98-100

[4]Noll A (1967) *Cepstrum pitch determination*. Journal of the Acoustical Society of America. 1967:41:293-309

[5] C.J. Moore, N. Slevin, and S. Winstanley, (1999) "Characterising vowel phonation by fundamental spectral normalisation of LX-waveforms", *Proceeding of the International Workshop on Models and Analysis of Vocal Emissions for Biomedical Applications*.1:1-6

[6]Kasuya H, Ogawa S, Mashima K, Ebihara S (1986) *Normalised noise energy as an acoustic measure to evaluate pathologic voice*. Journal of the Acoustic Society of America. 80(5):1329-1334

[7]McGillion (2000). *Automated Analysis of Voice Quality*. Ph.D. Thesis, UMIST